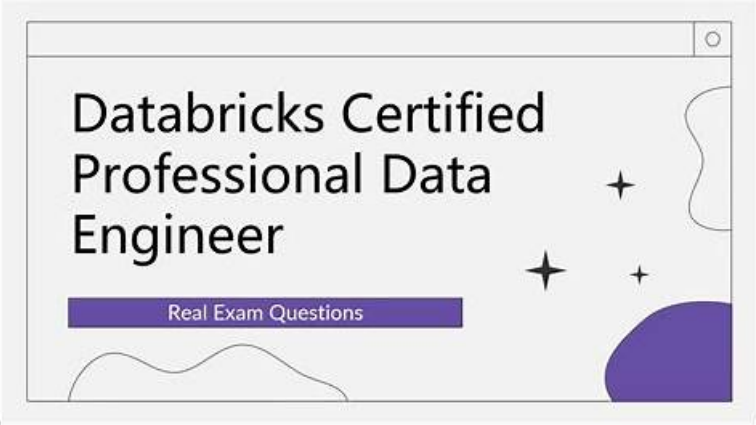
Valid Databricks-Certified-Professional-Data-Engineer Test Review - Latest Databricks-Certified-Professional-Data-Engineer Exam Question



There are three different Databricks Databricks-Certified-Professional-Data-Engineer questions format that is being provided to applicants from PrepAwayETE. Anyone can download a free Databricks-Certified-Professional-Data-Engineer exam dumps demo to evaluate this product before shopping. These Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) latest questions formats are Databricks Databricks-Certified-Professional-Data-Engineer PDF dumps format, web-based Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) practice tests, and desktop-based Databricks Databricks-Certified-Professional-Data-Engineer practice test software is provided to customers.

Databricks Certified Professional Data Engineer Certification Exam is created to challenge data engineers with the significant knowledge of Databricks' data engineering principles and techniques. To become Databricks certified, a candidate must pass the online certification exam designed for data engineers. Databricks-Certified-Professional-Data-Engineer Exam is scenario-based, comprises of 80 multiple-choice questions, and has a time limit of 120 minutes. The Certification exam tests the candidate's knowledge in topics such as data ingestion, data processing, data engineering, ETL, and data warehousing.

>> Valid Databricks-Certified-Professional-Data-Engineer Test Review <<

Free PDF Quiz Databricks-Certified-Professional-Data-Engineer - Latest Valid Databricks Certified Professional Data Engineer Exam Test Review

To assist applicants preparing for the Databricks Certified Professional Data Engineer Exam (Databricks-Certified-Professional-Data-Engineer) real certification exam effectively, PrepAwayETE offers Databricks Databricks-Certified-Professional-Data-Engineer desktop practice test software and a web-based practice exam besides actual PDF Databricks-Certified-Professional-Data-Engineer exam questions. These Databricks-Certified-Professional-Data-Engineer Practice Exams replicate the Databricks Databricks-Certified-Professional-Data-Engineer real exam scenario and offer a trusted evaluation of your preparation. No internet connection is necessary to use the Databricks-Certified-Professional-Data-Engineer Windows-based practice test software.

Databricks Certified Professional Data Engineer (Databricks-Certified-Professional-Data-Engineer) Certification Exam is a highly respected credential within the data engineering industry. Databricks Certified Professional Data Engineer Exam certification is specifically designed for professionals who have a deep understanding of data engineering principles, practices, and technologies. With this certification, data engineers can demonstrate their expertise in designing and building data pipelines, managing data workflows, and implementing data analytics solutions using Databricks.

Databricks Certified Professional Data Engineer Exam Sample Questions (Q198-Q203):

NEW QUESTION # 198
Data engineering team has a job currently setup to run a task load data into a reporting table every day at 8: 00 AM takes about 20

mins, Operations teams are planning to use that data to run a second job, so they access latest complete set of data. What is the best to way to orchestrate this job setup?

- A. Setup a Delta live to table based on the first table, set the job to run in continuous mode
- B. Setup a second job to run at 8:20 AM in the same workspace
- C. Use Auto Loader to run every 20 mins to read the initial table and set the trigger to once and create a second job
- D. Add Operation reporting task in the same job and set the Data Engineering task to de-pend on Operations reporting task
- E. Add Operation reporting task in the same job and set the operations reporting task to depend on Data Engineering task

Answer: E

Explanation:
Explanation
The answer is Add Operation reporting task in the same job and set the operations reporting task to depend on Data Engineering task.

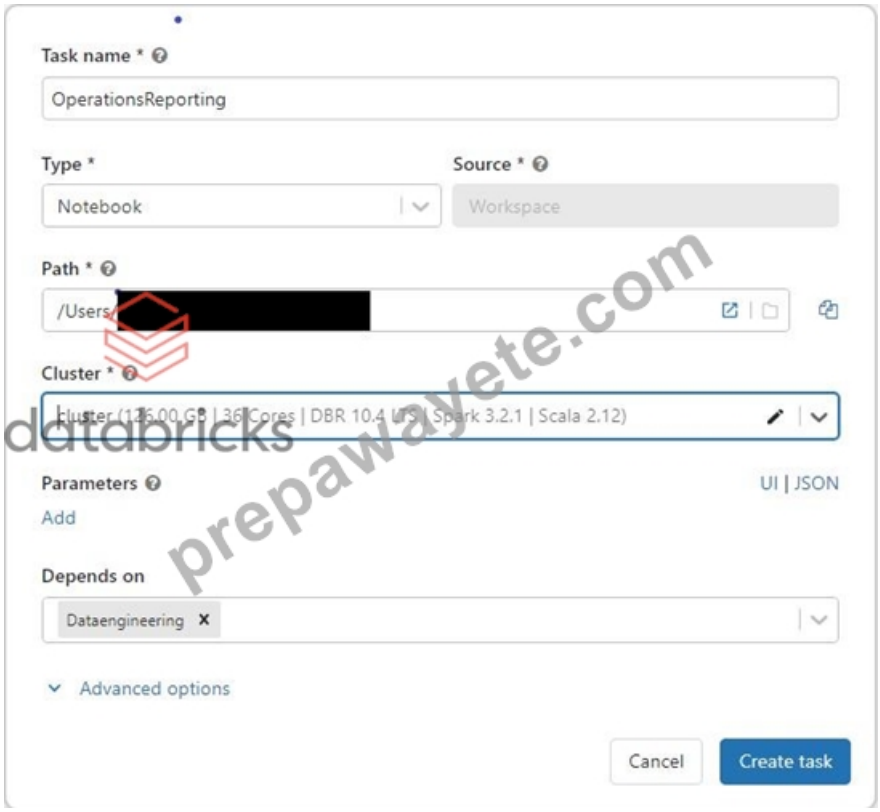
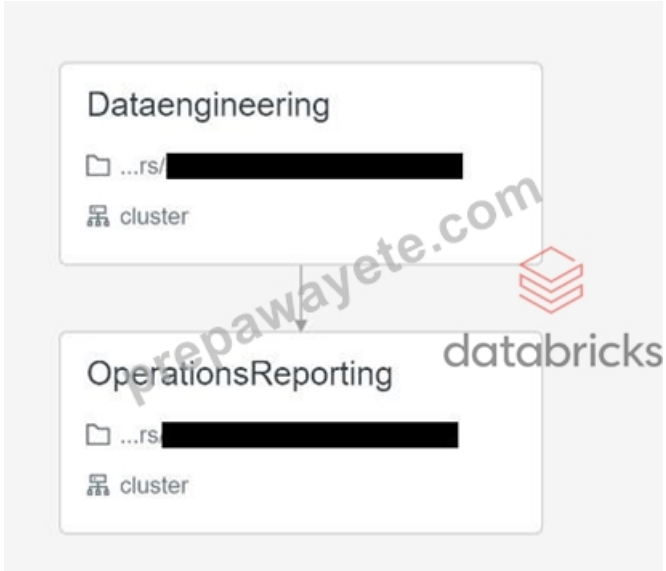


Diagram Description automatically generated with medium confidence



NEW QUESTION # 199
A Databricks job has been configured with 3 tasks, each of which is a Databricks notebook. Task A does not depend on other tasks. Tasks B and C run in parallel, with each having a serial dependency on task A.
If tasks A and B complete successfully but task C fails during a scheduled run, which statement describes the resulting state?

- A. Because all tasks are managed as a dependency graph, no changes will be committed to the Lakehouse until ail tasks have successfully been completed.
- B. Unless all tasks complete successfully, no changes will be committed to the Lakehouse; because task C failed, all commits will be rolled back automatically.
- C. All logic expressed in the notebook associated with tasks A and B will have been successfully completed; some operations in task C may have completed successfully.
- D. All logic expressed in the notebook associated with task A will have been successfully completed; tasks B and C will not commit any changes because of stage failure.
- E. All logic expressed in the notebook associated with tasks A and B will have been successfully completed; any changes made in task C will be rolled back due to task failure.

Answer: C

Explanation:
Explanation
The query uses the CREATE TABLE USING DELTA syntax to create a Delta Lake table from an existing Parquet file stored in DBFS. The query also uses the LOCATION keyword to specify the path to the Parquet file as /mnt/finance_eda_bucket/tx_sales.parquet. By using the LOCATION keyword, the query creates an external table, which is a table that is stored outside of the default warehouse directory and whose metadata is not managed by Databricks. An external table can be created from an existing directory in a cloud storage system, such as DBFS or S3, that contains data files in a supported format, such as Parquet or CSV.
The resulting state after running the second command is that an external table will be created in the storage container mounted to /mnt/finance_eda_bucket with the new name prod.sales_by_store. The command will not change any data or move any files in the storage container; it will only update the table reference in the metastore and create a new Delta transaction log for the renamed table. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "ALTER TABLE RENAME TO" section; Databricks Documentation, under "Create an external table" section.

NEW QUESTION # 200
The downstream consumers of a Delta Lake table have been complaining about data quality issues impacting performance in their applications. Specifically, they have complained that invalidlatitudeandlongitudevalues in theactivity_detailstable have been breaking their ability to use other geolocation processes.
A junior engineer has written the following code to addCHECKconstraints to the Delta Lake table:

A senior engineer has confirmed the above logic is correct and the valid ranges for latitude and longitude are provided, but the code fails when executed.
Which statement explains the cause of this failure?

- A. The activity details table already contains records; CHECK constraints can only be added prior to inserting values into a table.
- B. The current table schema does not contain the field valid coordinates; schema evolution will need to be enabled before altering the table to add a constraint.
- C. The activity details table already contains records that violate the constraints; all existing data must pass CHECK

constraints in order to add them to an existing table.

- D. The activity details table already exists; CHECK constraints can only be added during initial table creation.
- E. Because another team uses this table to support a frequently running application, two-phase locking is preventing the operation from committing.

Answer: C

Explanation:
Explanation
The failure is that the code to add CHECK constraints to the Delta Lake table fails when executed. The code uses ALTER TABLE ADD CONSTRAINT commands to add two CHECK constraints to a table named activity_details. The first constraint checks if the latitude value is between -90 and 90, and the second constraint checks if the longitude value is between -180 and 180. The cause of this failure is that the activity_details table already contains records that violate these constraints, meaning that they have invalid latitude or longitude values outside of these ranges. When adding CHECK constraints to an existing table, Delta Lake verifies that all existing data satisfies the constraints before adding them to the table. If any record violates the constraints, Delta Lake throws an exception and aborts the operation. Verified References:
[Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "Add a CHECK constraint to an existing table" section.

NEW QUESTION # 201
A Databricks job has been configured with 3 tasks, each of which is a Databricks notebook. Task A does not depend on other tasks. Tasks B and C run in parallel, with each having a serial dependency on task A.
If tasks A and B complete successfully but task C fails during a scheduled run, which statement describes the resulting state?

- A. Because all tasks are managed as a dependency graph, no changes will be committed to the Lakehouse until ail tasks have successfully been completed.
- B. Unless all tasks complete successfully, no changes will be committed to the Lakehouse; because task C failed, all commits will be rolled back automatically.
- C. All logic expressed in the notebook associated with tasks A and B will have been successfully completed; some operations in task C may have completed successfully.
- D. All logic expressed in the notebook associated with task A will have been successfully completed; tasks B and C will not commit any changes because of stage failure.
- E. All logic expressed in the notebook associated with tasks A and B will have been successfully completed; any changes made in task C will be rolled back due to task failure.

Answer: C

Explanation:
The query uses the CREATE TABLE USING DELTA syntax to create a Delta Lake table from an existing Parquet file stored in DBFS. The query also uses the LOCATION keyword to specify the path to the Parquet file as /mnt/finance_eda_bucket/tx_sales.parquet. By using the LOCATION keyword, the query creates an external table, which is a table that is stored outside of the default warehouse directory and whose metadata is not managed by Databricks. An external table can be created from an existing directory in a cloud storage system, such as DBFS or S3, that contains data files in a supported format, such as Parquet or CSV.
The resulting state after running the second command is that an external table will be created in the storage container mounted to /mnt/finance_eda_bucket with the new name prod.sales_by_store. The command will not change any data or move any files in the storage container; it will only update the table reference in the metastore and create a new Delta transaction log for the renamed table. Verified References: [Databricks Certified Data Engineer Professional], under "Delta Lake" section; Databricks Documentation, under "ALTER TABLE RENAME TO" section; Databricks Documentation, under "Create an external table" section.

NEW QUESTION # 202
The following code has been migrated to a Databricks notebook from a legacy workload:

```
%sh
git clone https://github.com/foo/data_loader;
python ./data_loader/run.py;
mv ./output /dbfs/mnt/new_data
```

The code executes successfully and provides the logically correct results, however, it takes over 20 minutes to extract and load

around 1 GB of data.
Which statement is a possible explanation for this behavior?

- A. Instead of cloning, the code should use %sh pip install so that the Python code can get executed in parallel across all nodes in a cluster.
- B. %sh executes shell code on the driver node. The code does not take advantage of the worker nodes or Databricks optimized Spark.
- C. %sh does not distribute file moving operations; the final line of code should be updated to use %fs instead.
- D. Python will always execute slower than Scala on Databricks. The run.py script should be refactored to Scala.
- E. %sh triggers a cluster restart to collect and install Git. Most of the latency is related to cluster startup time.

Answer: B

Explanation:
https://www.databricks.com/blog/2020/08/31/introducing-the-databricks-web-terminal.html The code is using %sh to execute shell code on the driver node. This means that the code is not taking advantage of the worker nodes or Databricks optimized Spark. This is why the code is taking longer to execute. A better approach would be to use Databricks libraries and APIs to read and write data from Git and DBFS, and to leverage the parallelism and performance of Spark. For example, you can use the Databricks Connect feature to run your Python code on a remote Databricks cluster, or you can use the Spark Git Connector to read data from Git repositories as Spark DataFrames.

NEW QUESTION # 203
.....

Latest Databricks-Certified-Professional-Data-Engineer Exam Question:
https://www.prepawayete.com/Databricks/Databricks-Certified-Professional-Data-Engineer-practice-exam-dumps.html

□Databricks-Certified-Professional-Data-Engineer Valid Exam Vce Free

- Databricks-Certified-Professional-Data-Engineer Real Exam Answers □ Databricks-Certified-Professional-Data-Engineer Reliable Test Sims □ Exam Databricks-Certified-Professional-Data-Engineer Book □ Easily obtain 【 Databricks-Certified-Professional-Data-Engineer 】 for free download through [www.pdfvce.com] □Databricks-Certified-Professional-Data-Engineer Practice Exam Fee
- Realistic Databricks Valid Databricks-Certified-Professional-Data-Engineer Test Review ♣ Search on ▸ www.prepawayexam.com ◂ for ➽ Databricks-Certified-Professional-Data-Engineer □ to obtain exam materials for free download □Databricks-Certified-Professional-Data-Engineer Valid Exam Vce Free
- petreligacademy.com, www.stes.tyc.edu.tw, myspace.com, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, hker2uk.com, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, Disposable vapes